# Development of Multifactor Dynamic Model of Forecasting of Scientific Migration

Anastasia E. Sudakova[1, a)], Andrey A. Tarasyev[1, b)]

[1]*Ural Federal University, 620002, Mira Street, 19, Ekaterinburg, Russia*

[a)]Corresponding author: a-chusova@mail.ru
[b)]andrej.tarasyev@urfu.ru

**Abstract.** The quality of human capital plays the key role in the technological and economic development of modern societies. On the current stage, many countries require highly skilled specialists due to a razing intensity of knowledge in technology. This is explained by the fact that highly skilled specialists contribute to the acceleration of knowledge accumulation, innovation implementation, and lead to the development of socio-economic development due to a higher level of economic activity. Therefore, to provide the country's development and competitiveness at the global level, it is important to maintain the academic exchange and to attract highly skilled specialists. Russia has a number of peculiarities in the migration process of highly skilled specialists. On the one hand, the country is experiencing the consequences of mass emigration of highly skilled specialists in the 20th century; on the other hand, it has a low ability to attract qualified specialists from other countries due to the specifics of the development of the socio-economic system. Features of socio-economic development of the country lead to the fact that currently in Russia obtain a unidirectional nature of the migration process of highly qualified personnel. Migration on a permanent basis and the mobility of the scientist has a different impact on the domestic science, on the increment of its intellectual capital. To solve effectively the problem of scientific emigration, it is necessary to assess the potential of scientific emigration and to be able to assess and predict the scale and direction of potential flows of scientific migration. The existing models and model complexes take into account macroeconomic dependencies to the detriment of the description of individual behavior, which limits the applicability of models to describe migration in a changing market environment. Due to the lack of a model complex that allows describing the variable dynamics of scientific migration between the labor markets at the macro level and scientific organizations at the micro level, it is necessary to analyze and evaluate methods for forecasting specific cases of labor resources redistribution between organizations, and to structure the main forecast methods. As a result, we developed a multivariate dynamic model of forecasting scientific migration with elements of the theory of positional games.
**Keywords:** behavioral economics, scientific migration, brain drain, dynamic model, data mining, Big Data, bibliometric assessment, optimization.

## INTRODUCTION

The most successful and dynamic models of economic and social development are based on the progress of knowledge-intensive industries. Therefore, huge investments are made in the development of fundamental and applied science. Despite the high cost of scientific infrastructure, one of the scarcest resources for the accelerated development of science is human resources and their potential. In these circumstances, the problem of scientific migration cannot be considered isolated from the global competition of scientists. Competition for talent influences innovation policy initiatives around the world [9]. For Russian science, which has, rather limited resources, in comparison with global trends, it is important to develop a strategy to respond to emerging challenges.

The problem of attracting and retaining competitive workers engaged in research and development, creating incentives for their work in domestic enterprises and organizations, as well as conditions for cooperation with foreign scientific organizations is a complex theoretical and practical task. Its solution is a fundamental task of assessing and predicting the dynamics, volume and direction of the migration movement of scientists.

*International Conference of Numerical Analysis and Applied Mathematics ICNAAM 2019*
AIP Conf. Proc. 2293, 120009-1–120009-4; https://doi.org/10.1063/5.0027280
Published by AIP Publishing. 978-0-7354-4025-8/$30.00

It is worth noting, that almost 60% of postdocs currently working in the US do so on a foreign visa [7]. Similarly, in Europe, 43% of doctoral students work in a country that is not their place of birth [1]. Despite the fact that the United States is considered the leading country for foreign researchers, whose share is 38.4%, the leading position in this ranking is Switzerland – 56.7%, Canada – 46.9% and Australia – 44.5%. In Europe, the following countries can be identified: UK-32.9%, Germany-23.2%, France-17.3%. At the end of the list are countries such as Japan – 5% of foreign researchers, Italy – 3%, India – 0.8%. The explanation for the low value of the indicator in each country is different. In Japan, it is due to the linguistic barrier. For Italy, it is believed that this is due to a lack of resources combined with a bad reputation caused by the phenomenon of nepotism. In India, a small number of foreign scientists involved are associated with cultural factors of the country and low wages in international comparison [3].

The expansion of the role of knowledge-based industries for economic development has increased scientific interest in the problem of highly qualified and scientific migration. In particular, extensive opportunities to attract scientists through return immigration are recognized as a competitive advantage of China's high-tech industries [5]. Looking at the hallmarks of modern globalization and techno capitalism, scientists [8] note, that technological creativity, corporate research, and talent flows are becoming more important than ever. They become the contextual central features of the globalization of the macro-social dynamics of the twenty-first century. Scientific migration cannot be considered in isolation from technology transfer. In this regard, the results obtained in the study of scientific mobility are of interest, indicating that organizations, whose employees are involved in the circulation of intelligence between countries, are the most productive in the creation of new technologies [2].

New theoretical approaches and concepts are developed to assess the processes of scientific migration. Many of them are debatable. For example, according to some experts [4], the modern model of migration of scientists is characterized by the transition from the model of brain drain (Brain drain) from one country and the growth of intellectual potential in another (Brain gain), to the model of intellectual potential exchange (brain sharing). One of the conclusions about the transition to the brain sharing migration model is the following thesis: international mobility benefits all parties, including countries that are net exporters of researchers.

Scientists expressed concern that the scientific team should have a certain number of researchers with the appropriate level of qualification. If team members begin the migration process, it will eventually lead to a General emigration of team members; this effect is called the Schelling's "dying seminar". Despite this theory, the prevention of technology leakage and the promotion of return emigration are the most important areas of research of practical importance for countries of origin in the field of migration of scientists [6].

## BIBLIOMETRIC ASSESSMENT

Given the weak structure and the presence of a significant share of the unstructured component of the analyzed data, the main data warehouse is the non-relational MongoDB database, as well as the Hadoop server based on Cloudera CDH. The main tool for implementing distributed data processing is Spark, which is interacted with through specialized packages of the R scripting language. The analysis of scientific mobility is carried out by the program of search and data processing, created on the basis of original algorithms developed by the authors of the project. The initial data array for articles is presented in the form of a column-matrix $A_{s1}=(a_{s1})_{r\times 1}, s=1, \ldots, r$, where the column is the considered organization, $s$ is the line with the article title, $r$ is the total number of articles with the analyzed affiliation of the organization. These matrix arrays are collected automatically for each year. At the second stage, the $A_{s1}$ matrix is expanded by adding new columns of characteristics to the analyzed data set: author ID, name, country of affiliation, additional affiliations, and the number of citations for the analyzed article. As a result, 7 data sets were obtained, reflecting Ural Federal University's academic mobility dynamics from 2011 to 2017.

To identify migrant scientists, a new matrix is being formed, reflecting information on how many works the authors have published with external organizations over the years. For this, we introduce $i$ - the parameter that determines the author's ID. Then, when the author search algorithm is launched in the assembled data array, the $B^k$ matrix will be formed in which the element $b_{ij}^k$ is obtained by accumulating values in case the authors' identification numbers in the array coincide with the authors' target numbers.

As a result of the transformations, the $B^k=(b_{ij}^k)_{m\times l}$ matrix is obtained, with the matrix lines representing all the authors of research papers, and the matrix columns ($i=1,\ldots,m$) - all the academic organizations identified. This matrix is compiled for each analyzed year, so each article has 3 main indexes that determine its position in the analyzed data array. It should be noted that the resulting matrix $B^k$ is a sparse matrix with elements reflecting information on the number of articles $b_{ij}^k$ of author $i$ from organization $j$ for selected year $k$.

In the compiled matrix, the first column represents the analyzed academic organization. If the affiliation of the main organization is identified, element $b_{ij}^k$ is fixed in the first column specifying the number of articles. When a change in the author's affiliation is detected, element $b_{ij}^k$ is assigned the value of the articles, and then it is written in column *j*, where *j* is the ordinal number of the column denoting the organization the author moved to. To output data on brain drain based on the obtained data array, the elements of matrix $B^k$ need to be checked for compliance with a number of conditions for time interval $k=2011, ..., 2017=1, ..., N$.

In general, the solution of the problem can be represented by a conditional breakdown of the total time into two parts: $k=1, ..., N/2$ and $k=N/2+1, ..., N$ and by putting the data on affiliation with the considered university in the first column $j=1$. As a result, the problem is reduced to the compilation of a final matrix $X=(x_{ij})_{c\times d}$ reflecting the number of publications affiliated with scientific organizations. The search and accumulation of the array of articles affiliated to authors *i* and academic organizations *j* on time interval $k=1, ..., N/2$, are performed as follows: $\sum_{k=1}^{N/2} b_{ij}^k = x_{ij}$, if $x_{ij}>3$, then author *i* is affiliated with organization *j*. The search on time interval $k=N/2+1, ..., N$ is performed in a similar way.

The result of this algorithm is a matrix of scientists indicating the main individual characteristics, including the direction of work, the number and quality of scientific works, as well as the current and previous places of work of the scientist. After the collection of statistical data, it is necessary to carry out its cleaning (removal and verification of homonyms, verification of authors who have a single publication and other parameters). The final matrix $X=(x_{ij})_{c\times d}$ reflects the number of articles written by author $i=1,...,c$ from academic organization $j=1,...,d$.

## DYNAMIC GAME MODEL OF HIGHLY SKILLED MIGRATION

The multifactor dynamic model of forecasting of highly skilled migration with elements of the theory of games is developed based on provisions of the theory of positional games taking into account accomplishment of conditions of equilibrium on Nash for the labor market. The model allows us to track and predict the migration of scientists depending on the availability of vacancies, wages and the degree of development of socio-economic systems of regions and countries in which the centers of attraction of scientific migration are located, which corresponds to the model of brain drain (brain drain and brain gain). The proposed model is used to predict migration processes that depend on the level of wages and the degree of development of socio-economic systems of the regions and countries in which the centers of attraction of scientific migration are located. Thus, when changing the structure of the object of attraction in the course of model calculations, migration flows can be reoriented to the newly appeared in the simulation centers of attraction. The dynamics of migration flows between scientific organizations is described by the following function:

$$y_{ij}(t+1) = y_{ij}(t) + a_j \cdot y_{ij}(t) \cdot \frac{\left(P_i - y_{ij}(t)\right)}{P_i} \cdot \frac{\left(s_n(t) - s_m(t) - C_{mn}\right)}{s(t)} dt \, , \tag{1}$$

where $s_m(t)$, $s_n(t)$ – denotes the price of labor, the difference in wages levels in organizations involved in the migration process; $P_i$ denotes the potential amount of scientists, ready for migration; $y_{mn}(t)$ denotes the migration flow from the organization *m* recalculated from matrix $X=(x_{ij})_{c\times d}$; $C_{mn}$ denotes the costs of migration, $C_{mn} = M_{mn}/N$.

Here the dynamics of scientific migration depends on the main model driver, which is based on the salaries gap and migration costs of moving. On the first step of the analysis we estimate the share of scientists, ready to migration, based on the data obtain from the mining algorithm. These data include the information on scientist's mobility, the amount of published scientific papers, the share of papers with affiliation to several scientific organizations and field of study for each scientist. On the second step of the conducted analysis, we estimate the possible benefits for each scientist. If the expected salary in the organization of scientist's attraction overlap the migration costs and the possible salary in the organization of scientist's origin, $s_n(t)-s_m(t)-C_{mn}>0$, than the decision of migration will be positive. In the other case, if $s_n(t)-s_m(t)-C_{mn}<0$, the decision will be negative. The third possible decision, which includes the further migration of a scientist is based on the condition, that he can obtain more benefits from third organization *K*, where $s_K(t)-s_m(t)-C_{mn}>0$.

The average scientist's salaries level depends on the amount of academic staff in the organization, their qualification level and their competition for higher salary. The estimation of academic staff in the organization is obtained using the amount of scientific works written with affiliation to the current organization. In this case, we estimate also the amount of scientific capital in each scientific organization. The wages function for the organization of scientist's origin will be described using function (2):

$$s_m(t) = s_m(t-1) \cdot \left( \frac{E_m(t)}{E_m(t) - y_{mn}(t)} + \frac{x_{im}(t)}{x_m(t)} \right). \tag{2}$$

The wage function for the organization of scientific migration attraction is structured similarly to (2), allows obtaining a form of Nash equilibrium in the model at the infinite horizon:

$$s_n(t) = s_n(t-1) \cdot \left( \frac{E_n(t)}{E_n(t) + y_{mn}(t)} + \frac{x_{in}(t)}{x_n(t)} \right). \tag{3}$$

## CONCLUSION

The multifactor dynamic model of forecasting migration of highly qualified personnel with elements of the game theory is developed by authors based on construction of the theory of positional games taking into account realization of equilibrium conditions of Nash for the labor market. This model makes it possible to track and predict the migration of scientists depending on the availability of vacancies, the level of wages and the level of development of socio-economic systems of the regions and countries in which the centers of attraction of scientific migration are located. This question follows the model of "brain drain" (brain drain and brain gain). Model dynamics is based on the theory of human capital, behavioral Economics, new migration Economics, neoclassical migration theory and synthetic migration theory. As a result, a model of the dynamics of scientific migration flows between large centers of attraction, based on the principle of the potential difference relative to the basic parameters of the movement (drivers) of the model. The proposed model is used to predict migration processes depending on the level of wages and the degree of development of regional socio-economic systems, which are located in the centers of attraction of scientific migration. Thus, when changing the structure of the object of attraction in the course of model calculations, migration flows are reoriented to the newly emerging centers of attraction in the modeling process.

## ACKNOWLEDGEMENTS

Research is supported by the project of the Russian Science Foundation No. 19-78-00080, “Investigation of migration of Russian scientists on the basis of predictive modeling and analysis of data on Big Data and Data Science technologies”.

## REFERENCES

[1] Barjak, F. (2005). Doctoral Students in the Life Sciences. Bonn, Germany: Institute for Prospective Technological Studies, Commission of the European Communities. Empirica GmbH.
[2] Edler, J., Fier, H., Grimpe, C. (2011). International scientist mobility and the locus of knowledge and technology transfer. Research Policy. Vol. 40. Is. 6. pp. 791-805.
[3] Franzoni, C., Scellato, G., Stephan, P. (2012). Foreign-born scientists: mobility patterns for 16 countries. Nat Biotechnol 30. pp. 1250–1253.
[4] Hunter, P. (2013). Brain drain, brain gain or brain sharing? New studies of the migration routes of scientists show that international mobility benefits all parties including countries that are net exporters of researchers. EMBO Reports. Vol. 14. Is. 4. pp. 315-318.
[5] Kshetri, N. (2008). The Rapidly Transforming Chinese High-Technology Industry and Market: Institutions, Ingredients, Mechanisms and Modus Operandi. UK, Oxford: Chandos Publishing. 328 p.
[6] Saint-Blancat, C. (2017). Making Sense of Scientific Mobility: How Italian Scientists Look Back on Their Trajectories of Mobility in the EU. Higher Education Policy.
[7] Stephan, P. (2012). How Economics Shapes Science. Cambridge, Massachusetts: Harvard University Press. 384 p.
[8] Suarez-Villa, L. (2012). Globalization and Technocapitalism: The Political Economy of Corporate Power and Technological Domination. UK: Ashgate Publishing Ltd. 243 p.
[9] The Global Competition for Talent: Mobility of the Highly Skilled. Paris: OECD Publishing. 2008. 165 p.